

Review

## Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology

Marina Barba <sup>1,\*</sup>, Henryk Czosnek <sup>2</sup> and Ahmed Hadidi <sup>3,†</sup>

<sup>1</sup> Consiglio per la ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per la Patologia Vegetale, Via C. G. Bertero 22, Rome 00156, Italy

<sup>2</sup> Institute of Plant Sciences and Genetics in Agriculture, The Robert H. Smith Faculty of Agriculture Food and Environment, The Hebrew University of Jerusalem, Rehovot 76100, Israel;  
E-Mail: hanokh.czosnek@mail.huji.ac.il

<sup>3</sup> U.S. Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705, USA;  
E-Mail: ahadidi@yahoo.com

† Emeritus Lead Scientist.

\* Author to whom correspondence should be addressed; E-Mail: marina.barba@entecra.it;  
Tel.: +39-06-8207-0244; Fax: +39-06-8207-0243.

Received: 15 October 2013; in revised form: 17 December 2013 / Accepted: 24 December 2013 /  
Published: 6 January 2014

---

**Abstract:** Next-generation high throughput sequencing technologies became available at the onset of the 21st century. They provide a highly efficient, rapid, and low cost DNA sequencing platform beyond the reach of the standard and traditional DNA sequencing technologies developed in the late 1970s. They are continually improved to become faster, more efficient and cheaper. They have been used in many fields of biology since 2004. In 2009, next-generation sequencing (NGS) technologies began to be applied to several areas of plant virology including virus/viroid genome sequencing, discovery and detection, ecology and epidemiology, replication and transcription. Identification and characterization of known and unknown viruses and/or viroids in infected plants are currently among the most successful applications of these technologies. It is expected that NGS will play very significant roles in many research and non-research areas of plant virology.

**Keywords:** next-generation (deep) sequencing; NGS; novel virus/viroid discovery; metagenomics; virome; transcriptome; DNA sequencing; RNA sequencing (RNA-seq.)

---

## 1. Introduction

Next-generation (high throughput, deep) sequencing (NGS) has been developed in recent years. These technologies have lowered the costs of DNA sequencing beyond what is possible with standard dye-terminator methods. NGS describes platforms that produce large amounts (typically millions to billions) of DNA reads, with lengths between 25 and 400 bp. These reads are shorter than the traditional Sanger sequence reads (300 to 750 bp). Recently, however, NGS technologies have advanced to produce DNA reads longer than 750 bp. The roots of NGS development go back to the discovery of the structure of the DNA double helix in 1953 by James Watson and Francis Crick, an American geneticist and a British physicist, respectively [1]. Robert Holley, an American biochemist, was the first to sequence a nucleic acid when he and colleagues developed sequencing methods for tRNA in 1964 and 1965 [2,3]. Sequencing methods for long DNA were successfully developed independently in 1977 by Frederick Sanger, a British biochemist [4,5], and by Walter Gilbert, an American biochemist/physicist [6]. The Nobel Prize was received by Watson and Crick in 1962, by Holley in 1968, and by Sanger and Gilbert in 1980.

## 2. Pioneer Landmarks in DNA Sequencing

Several years before any attempt was made to sequence a DNA molecule, in 1964 Robert Holley (1922–1993) was the first to sequence an RNA molecule. He determined the complete sequence and structure of the 77 ribonucleotides of alanine tRNA, the molecule that incorporates the amino acid alanine into protein. Holley's pioneer work opened the door for others to determine the sequence of other RNAs as well as DNA.

In 1972, Paul Berg (1926–present) developed DNA technology, which permits isolation of defined fragments of DNA [7]. Prior to this, the only accessible samples for sequencing were from phages or virus DNA. Also, his discovery led to the development of the modern genetic engineering.

In 1973, the first nucleotide sequence of 24 bp out of 27 bp of the *lac* operator DNA was published [8]. In 1977, Frederick Sanger (1918–2013) was the first to sequence the complete DNA genome of bacteriophage  $\Phi$ X 174 [4]. He also developed “DNA sequencing with chain-terminating inhibitor” [5]. Moreover, in 1977 Walter Gilbert (1932–present) developed “DNA sequencing by chemical degradation” [6].

Paul Berg, Frederick Sanger and Walter Gilbert received the Nobel Prize in Chemistry in 1980. Frederick Sanger received the Noble Prize twice, the first one in 1958 for his work on the structure of proteins, especially that of insulin. Sanger is one of only four people to win two Nobel Prizes and the only Nobel Laureate to win two chemistry prizes.

## 3. Major Landmarks in DNA Sequencing during the Last Three Decades

1984: Medical Research Council scientists [9] completed the DNA sequence of the *Epstein-Barr virus* (EBV) 172,282 bp using the dideoxynucleotide/M13 sequencing procedure. EBV causes infectious mononucleosis.

1986: LeRoy Hood at the California Institute of Technology (Cal Tech., Pasadena, CA, USA) announced the invention of the first-semi-automated DNA sequencing machine. The machine automated the enzymatic chain termination procedure for DNA sequence analysis developed by Sanger and became a key instrument in mapping and sequencing genetic material.

1987: Applied Biosystems (USA) marketed the first automated sequencing machine, the model ABI 370. Constant improvements in the technology resulted in faster sequencing capacity, which was significant for advanced scientific research in projects such as mapping the human genome.

1990: The International Project on Human Genome was formally started and it was expected to take 15 years. It involved mainly the United States, the United Kingdom, France, Germany, Japan, China, and India.

1990: The US National Institutes of Health (NIH) began large-scale sequencing trials on *Mycoplasma capricolum*, *Escherichia coli*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae*.

1995: Craig Venter, Hamilton Smith and colleagues (USA) completed the first complete genome of a free-living organism, the bacterium *Haemophilus influenzae*. The circular chromosome contains 1,830,137 bp [10]. This is the bacterium from which Smith had first isolated a restriction enzyme for which he shared the Nobel Prize in 1978.

1996: Applied Biosystems markets the capillary sequencer, ABI 310. It is an automatic single-capillary genetic analyzer designed for a wide range of sequencing and fragment analysis approaches.

1996: Pal Nyren and Mostafa Ronaghi (Sweden) developed the method of DNA pyrosequencing without the need for electrophoresis [11]. It relies on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay developed by P. Nyren in 1987.

1998: Eric Kawashima, Laurent Farinelli and Pascal Mayer developed “Method of nucleic acid amplification” and obtained WO 98/44151 patent for its development [12] while working at the Glaxo Wellcome’s, Geneva Biomedical Research Institute. This method described DNA colony sequencing which is one of the milestones in developing the massive parallel sequencing technologies such as Illumina (San Diego, CA, USA) and Life Sciences 545 (a Roche Company, Branford, CT, USA).

1998: Phil Green and Brent Ewing (USA) developed the base calling program “Phred” for sequencer data analysis. It has the ability to estimate a probability of error for each base-call, as a function of certain parameters computed from the tracer data [13].

1998: Genome of the nematode *Caenorhabditis elegans* sequenced. The 97-megabase pair genomic sequence revealed over 19,000 genes. The complete genome sequence, the first from a multicellular organism and from an animal, was a joint sequencing project between the Wellcome Trust Sanger Institute (Hinxton, UK) and the Genome Institute, Washington University, St. Louis, MO, USA.

2000: Human Genome Project: due to widespread international cooperation and advances in the field of genomics (especially in sequence analysis), as well as major advances in computing technology, a “rough draft” of the genome was finished in 2000 (announced jointly by U.S. President Bill Clinton and the British Prime Minister Tony Blair on June 26, 2000).

2001-2003: Human Genome Project: Key findings of the draft (2001) and essentially complete genome were announced in April 2003, 2 years earlier than planned (3.3 billion-base pairs; approximately 23,000 genes).

#### 4. The Need for Fast, Inexpensive and Accurate DNA Sequencing Technologies

The automated Sanger method had led to a number of major accomplishments, including the completion of the human genome and other selected animal and plant genomes. However, the method limitations showed a need for new and improved technologies for sequencing large numbers of human and other genomes. In the late 20th and early 21st century, efforts have been made towards the development of new methods to replace the automated Sanger method, which is considered as a “first-generation” technology. The newer methods are referred to as next-generation sequencing (NGS) and their use has changed the scientific approaches in both basic and applied research in many of scientific disciplines, especially in many branches of the biological field, including plant pathology and plant virology.

The major advance offered by NGS is the ability to produce an enormous volume of data, in several cases in excess of one billion short reads per instrument run, as well as its ability to deliver fast, inexpensive and accurate genome information.

#### 5. Development of NGS Platforms (2000–present)

In 2000, Massively Parallel Signature Sequencing (MPSS) Lynx Therapeutics (USA) Company launched the first of the NGS technologies. The company was later purchased by Illumina.

In 2004, 454 Life Sciences (Branford, CT, USA) marketed a paralleled version of pyrosequencing. The first version of their machine reduced sequencing costs 6-fold compared to automated Sanger sequencing and was the second of a new generation of sequencing technologies, after MPSS. Life Sciences acquired by Roche Company (Headquartered in Basel, Switzerland). Pyrosequencing provides intermediate read lengths and price per base compared to Sanger sequencing on the one hand and Illumina and SOLiD on the other. In 2005–2006, the 454 GS 20 Roche sequencing platform was introduced which revolutionized DNA sequencing as it could produce 20 million bases (20 Mbp). This was replaced by the GS FLX model in 2007 which is capable of producing over 100 Mbp of sequence in just four hours, which increased in 2008 to 400 Mbp. This model was then upgraded to the 454 GS-FLX+ Titanium sequencing platform which is capable of producing over 600 Mbp of sequence data in a single run with Sanger-like read lengths of up to 1,000 bp. Another sequencing platform system produced by Roche is the GS Junior which is small in size and is placed on a laboratory bench top. It provides long 400 bp sequencing reads with a fast sequencing run [14].

In 2005, Solexa released the Genome Analyzer (GA). Its sequencing technology is based on sequencing by synthesis (SBS) using reversible dye-terminators chemistry. Solexa was purchased by Illumina in 2007. The GAIIx platform generates up to 50 billion bases (50 Bbp) of usable data per run and the latest model can attain 85 Bbp per run.

During the last 3–4 years, Illumina has developed the HiSeq platform series which include HiSeq® 2500, HiSeq 2000, HiSeq 1500 and HiSeq 1000 sequencing platforms. They vary in their outputs, run times, cluster generations, paired end reads, and maximum reads length with HiSeq 2500 the longest length of 200 bp and HiSeq 1000 the shortest one. HiSeq 2500 platform has the capacity to sequence a human genome in about 24 hours, “Genome in a day”. It may also sequence 20 exomes in a day, or 30 RNA sequencing samples in as little as five hours. It generates 120 billion b (120 Bbp) of data in 27 hours. Standard HiSeq 2000 generates 600 billion bases (600 Bbp) per run. The output of HiSeq

2500 can reach 600 Gbp. Illumina also released MiSeq in 2011, a bench top platform which shared most technologies with HiSeq series. It generates 1.5 Gbp per run in about 10 h [15].

SOLiD technology employs sequencing by oligo ligation detection. The result is sequences of quantities and length comparable to Illumina sequencing. SOLiD was purchased by Applied Biosystems (AB, founded in 2006 in Foster City, CA, USA and became a division of Life Technologies in 2008 Headquartered in Carlsbad, CA, USA). Owing to a two-base sequencing method, SOLiD accuracy may reach 99.99%.

In 2007, the first SOLiD sequencing system was released by AB, followed by the SOLiD 5500 w and 5500 xlw sequencing systems in 2010. The SOLiD 5500 xlw has read lengths of 85 bp, with 99.99% accuracy and 30 Gbp per run. A complete run could be finished within a week. Applications of SOLiD include analysis of whole genome clusters. The SOLiD website, see [16], may be helpful.

Other recently developed methods of NGS technologies include: Helicos sequencer released in 2009, Life Technologies Ion Torrent sequencer released in 2011 [17], Pacific Biosciences (Menlo Park, CA, USA) single molecule real-time (smrt) sequencer which also became available in 2011 [18], and Oxford Technologies Nanopore (Oxford, UK) single molecule sequencer with ultra long single molecule reads that became available in 2012–2013 [19]. In November 2012, Helicos Biosciences (Cambridge, MA, USA) filed for bankruptcy in Cambridge, MA, USA, and it is currently reorganizing its operation. For this reason its website is no longer available.

Additional methods of NGS were also developed. These methods may include Polonator sequencing, Polony sequencing, DNA Nanoball sequencing, and VisiGen Biotechnologies sequencing.

Table 1 shows current major methods of NGS technologies and Table 2 shows current models of major sequencing platform systems and their applications in NGS technologies.

## 6. Platform Selection

Important factors in selecting a sequencing platform may include the size or expected size of the genome being studied, its complexity (including G+C content), and the depth of coverage and accuracy needed. Thus, it is important and advisable to contact the providers of the next-generation sequencing services for guidance. For *de novo* genome sequencing, longer read length may be appropriate. For fast turnover times and limited throughput, smaller laboratory bench top platforms may offer greater flexibility [20]. For amplicon sequencing, the Roche 454 platform is suitable because of its longer reads, however, it is currently expensive. Recently, the laboratory bench tops Roche 454 GS Junior, Illumina MiSeq, and Ion PGM have claimed that their sequencing platforms are suitable for sequencing amplicons. For RNA-seq and those projects that require high depths of coverage, Illumina and SOLiD platforms may offer the best all-round value for money, accuracy and throughput [21]. Roche 454 has the longest read length, Illumina HiSeq 2500 features the biggest output and lowest sequencing cost, and SOLiD 5500 xlw has the highest accuracy [22]. Tables 3–8 show that during the last five years, Illumina sequencing platforms were used more frequently than those of Roche's in different plant virology projects in several countries.

**Table 1.** Current major methods of next-generation DNA sequencing technologies.

Sequencing platform	Amplification method	Sequencing chemistry	Read length (bp)	Sequencing Speed/h	Maximum Output Per run	Accuracy (%)	M <sup>1</sup> I <sup>2</sup> D <sup>3</sup>
454 (Roche)	Emulsion PCR	Pyrosequencing	400–700	13 Mbp	700 Mbp	99.9	0.10, 0.3, 0.02 [23]
Illumina (Illumina)	Bridge PCR	Reversible terminators	100–300	25 Mbp	600 Gbp	99.9	0.12, 0.004, 0.006 [23]
SOLiD (Life Technologies)	Emulsion PCR	Ligation	75–85	21–28 Mbp	80–360 Gbp	99.9	Error is higher than Illumina [24]
PacBio (Pacific Biosciences)	No amplification Single molecule real-time (or SMRT)	Fluorescently labeled nucleotides	4,000–5,000	50–115 Mbp	200 Mb–1 Gbp	95	1, 2, 12 [25]
Helicos (Helicos Biosciences)	No amplification Single molecule	Reversible terminators	25–55	83 Mbp	35 Gbp	97	Error is in the range of few percent but higher than 454 and Illumina and biased toward InDels [24]
Ion Torrent (Life Technologies)	Emulsion PCR	Detection of released H	100–400	25 Mb–16 Gbp	100 Mb–64 Gbp	99	M, 0.06, I + D 1.38 [26]
Nanopore (Oxford Technologies)	No amplification Single molecule		Very long reads up to 50 kbp	150 Mbp	Tens of Gbp	96	

M<sup>1</sup> = Mismatch bases; I<sup>2</sup> = Insertion; D<sup>3</sup> = Deletion.

**Table 2.** Current models of major sequencing platform systems and their applications in next-generation sequencing technologies.

Platform Systems	Applications
454 GS FLX + Systems (GS FLX Titanium XL+/GS FLX Titanium XLR70) * GS Junior System (bench top)	DNA sequencing: whole genome sequencing, <i>de novo</i> and resequencing of large genes in a single run with read length up to 1 kbp. Amplicon sequencing; RNA sequencing: transcriptome sequencing, sequencing capture metagenomics. Run time: 10–23 h; * Fast sequencing run with read length of 400 bp. Similar applications as the above system.
Illumina HiSeq Systems (2500/2000/1500/1000), Genome Analyzer IIX, HiScan SQ, * MiSeq (bench top)	DNA sequencing including candidate region targeted sequencing; Epigenetic sequencing: chromatin immunoprecipitation sequencing (ChIP-Seq), methylation analysis by sequencing; RNA sequencing: transcriptome analysis, small RNA and mRNA sequencing, gene expression profiling by sequencing; Run time: 8–14 days; * Up to 15 Gbp of output with 25 M sequencing reads and 2 × 300 bp read length; Access more sequencing applications such as exome, metagenomics, human leukocyte antigen (HLA) gene typing, mRNA sequencing, targeted gene expression (proteins and non-protein coding genes) such as rRNA, tRNA, or smRNA genes; Run time: 20–35 h.

Table 2. Cont.

Platform Systems	Applications
SOLiD 5500 W Series Genetic Analysis Systems (5500 W, 5500 xlw)	DNA sequencing: whole genome and exome; Epigenetic Sequencing; RNA Sequencing. Run time: 7–12 days.
PacBio PACBIO RSII	DNA sequencing using single molecule real-time (SMRT) system with the longest read lengths of any sequencing technology. Characterization of genetic variation, methylation, targeted sequencing such as SNP detection and validation, indels, structural variants, haplotypes and phasing, base modification detection to understand gene expression, host-pathogen interactions, DNA damage and DNA repair. Run time: 30 min.
Helicos Genetic Analyzer System	DNA sequencing and RNA sequencing. Run time: 8 days.
Ion Torrent Ion PGM System (bench top) *Ion Proton System (bench top)	Semiconductor sequencing with 400-bp length- Ideal for sequencing small genes and genomes. DNA sequencing for microbial: genes and genomes, amplicons, exomes (unreveal disease-causing variants), targeted sequencing, viral typing and other microbial typing. RNA sequencing. Run time: 4.5 h * Semiconductor sequencing: Sequencing microbial genomes, exomes, transcriptomes. Run time: 4.5 h
Nanopore GridION System (bench top) *MinION System (a miniaturized disposable device for single use)	DNA sequencing; Epigenetic sequencing; Characterization of genetic variation. RNA sequencing: the system is designed to analyze the original sample RNA directly, without undergoing conversion to cDNA. Run time possibly under 60 min. * For DNA sequencing only, <i>i.e.</i> , blood DNA.

**Table 3.** Next-generation sequencing of plant viral siRNA, RNA or DNA from virus infected herbaceous, grass, fiber- producing, or vine hosts.

Host	Study finding/virome	Sample preparation/target	Sequencing platform	Ref.
Sweet potato	Detected: <i>Sweet potato feathery mottle virus</i> , <i>Sweet potato chlorotic stunt virus</i> ; Discovered: two <i>Badnavirus</i> species (dsDNA), one <i>Mastrevirus</i> species (ssDNA)	siRNAs	Illumina	[27]
<i>Gomphrena globosa</i>	Plants were inoculated with unknown virus-a new <i>Cucumovirus</i> was identified. Proposed name: Gayfeather mild mottle virus	Total RNA + Subtractive hybridization	Roche 454 GS-FLX	[28]
<i>Arabidopsis thaliana</i>	<i>Tobacco mosaic virus</i> siRNAs mediate virus-host interactions which may contribute to viral pathogenicity and host specificity	siRNAs	Illumina Genome Analyzer	[29]
<i>Nicotiana benthamiana</i> , <i>Arabidopsis thaliana</i> , <i>Cucumis melo</i> and tomato	Nine different viruses including <i>Cucumber mosaic virus</i> , <i>Tobacco rattle virus</i> , <i>Pepper mild mosaic virus</i> , <i>Potato virus X</i> were studied. The study extended the knowledge of distribution and composition of siRNAs in virus-infected plants and contributed to a better understanding of siRNAs biogenesis.	siRNAs of nine different viruses	Roche 454	[30]
Cassava	The complete sequence of the Tanzanian strain of <i>Cassava brown streak virus</i> was determined and compared with that of the Ugandan strain. The virus is highly heterogeneous at both the isolate and strain levels with nucleotide identity at the isolate level of 76%	Total RNA + subtractive hybridization	Roche 454 GS-FLX	[31]
<i>Nicotiana benthamiana</i>	Profiled <i>Cymbidium ringspot virus</i> -derived siRNAs. These RNAs were primarily produced from the positive strand of the virus, produced with different frequency, and had 5' monophosphate and were not perfect duplexes.	siRNAs	Roche 454 and Solexa (Illumina)	[32]
Wild plant species from 15 families naturally infected with viruses were utilized. The families are: <i>Acanthaceae</i> , <i>Bignoniaceae</i> , <i>Caesalpinaceae</i> , <i>Commelinaceae</i> , <i>Cyperaceae</i> , <i>Cucurbitaceae</i> , <i>Euphorbiaceae</i> , <i>Gesneriaceae</i> , <i>Lamiaceae</i> , <i>Mimosaceae</i> , <i>Myrtaceae</i> , <i>Papilionaceae</i> , <i>Poaceae</i> , <i>Rubiaceae</i> and <i>Solanaceae</i>	Identification of 11 virus families in infected plants which include: <i>Bromoviridae</i> , <i>Caulimoviridae</i> , <i>Chrysoviridae</i> , <i>Closteroviridae</i> , <i>Endornaviridae</i> , <i>Luteoviridae</i> , <i>Narnaviridae</i> , <i>Partitiviridae</i> , <i>Potyviridae</i> , <i>Totiviridae</i> , and <i>Tymoviridae</i> . Unclassified virus families were also identified in some samples of infected plants. Discovered several thousand novel viruses, all linked to their specific plant hosts	dsRNAs	Roche 454- GS-FLX	[33]

Table 3. Cont.

Host	Study finding/virome	Sample preparation/target	Sequencing platform	Ref.
Wild cocksfoot grass	<i>Cereal yellow dwarf virus (Luteovirus)</i> discovered infecting this grass (not previously reported infecting this host). <i>Cocksfoot streak virus (Potyvirus)</i> was detected	siRNAs	Roche 454	[34]
<i>Nicotiana benthamiana</i> , <i>Arabidopsis thaliana</i>	Profiling siRNAs of <i>Bamboo mosaic virus</i> and its interfering and non-interfering associated satellite RNAs. The overall composition of virus siRNAs and satellite RNAs in the infected plants reflect the combined action of virus, satellite RNA and different DCLs in host plants.	siRNAs (virus and satellites)	Solexa (Illumina)	[35]
Rice	Characterization of siRNAs of <i>Rice stripe virus</i> four genome RNAs in infected rice plants	siRNAs	Illumina Solexa	[36]
Cotton	Characterization of siRNAs of <i>Cotton leafroll dwarf virus</i> (genus <i>Polerovirus</i> , family <i>Luteoviridae</i> ) in infected cotton plants	siRNAs	Illumina Genome Analyzer	[37]
Wild <i>Passiflora caerulea</i> (Blue-crown passion flower) (a vine)	The complete nucleotide sequence of <i>the Passion fruit woodiness virus (Potyvirus)</i> was determined	Polyadenylated RNA	Illumina Solexa GAIIx	[38]
Pepper, eggplant	The complete nucleotide sequences of two new viruses <i>Pepper yellow leaf curl virus (Polerovirus)</i> and <i>Eggplant mild leaf mottle virus (Ipomovirus)</i> were determined	Purified virions viral RNA	SOLiD	[39]
Different hosts	A novel virus infecting watercress was identified (proposed name: Watercress white vein virus). Viruses such as <i>Piper yellow mottle virus</i> , <i>Arachacha virus B</i> , and <i>Potato black ring virus</i> were detected and sequenced	Partial virus and RNA purification then fragmented	Roche 454 GS-FLX	[40]
Tomato	siRNAs of <i>Tomato spotted wilt virus</i> were used to detect the virus in infected tomato before symptoms appeared at levels too low for conventional detection methods. Also, used for analysis of the virus quasi species and for identification of an unspecified <i>Tospovirus</i> and a squash-infecting geminivirus	siRNAs	Illumina	[41]
<i>Arabidopsis thaliana</i>	Characterization of the siRNAs and transcriptome profiles of <i>Oilseed rape mosaic virus (Tobamovirus)</i> -infected <i>Arabidopsis</i> plants.	Total RNAs and siRNAs	Illumina Genome Analyzer	[42]

Table 3. Cont.

Host	Study finding/virome	Sample preparation/target	Sequencing platform	Ref.
Tomato, <i>Nicotiana benthamiana</i>	Characterization of the siRNAs for the monopartite begomovirus <i>Tomato yellow leaf curl China virus</i> and its associated betasatellite in infected tomato and <i>Nicotiana benthamiana</i> plants. Also, it was found that the betasatellite affected the amount of virus siRNAs detected in both plant species.	siRNAs	Solexa-Illumina	[43]
Tomato	Identification of <i>Potato spindle tuber virod</i> , <i>Pepino mosaic virus</i> , differentiation of two strains of the virus, and a novel <i>Potyvirus</i> from infected tomato plants by obtaining the complete genome sequence of the mixed infected pathogens without prior knowledge of their existence. Based on the severity of symptoms on tomato, the novel virus is provisionally named “Tomato necrotic stunt virus”	siRNAs	Illumina Genome Analyzer IIx	[44]
Tobacco cv Xanthi nc	Identification of gene expression changes associated with disease development in tobacco plants induced by infection with the M strain of <i>Cucumber mosaic virus</i> Sequencing analysis of tobacco transcriptome identified 95,916 unigenes, 34,408 of which were new transcripts by database searches.	Total RNA then treated with DNase I	Illumina Hi Seq 2000	[45]
Corn (maize)	Detection and identification of <i>Maize chlorotic mottle virus</i> and <i>Sugarcane mosaic virus</i> by obtaining over 90% of both viral genomes sequencing which allowed also strain characterization. Next generation sequencing may be used for rapidly identifying potential disease causing agents, in this case viruses	Total RNA was purified from diseased tissue with virus symptoms	Roche 454 GS-FLX+	[46]
Seventeen plant species, among them 14 Australian indigenous species	Detection and identification of 12 viruses described previously members of the genera <i>Potyvirus</i> , <i>Nepovirus</i> , <i>Allexivirus</i> and <i>Carlavirus</i> . Four novel viruses were identified and proposed as members of the genera <i>Potyvirus</i> , <i>Sadwavirus</i> and <i>Trichovirus</i> . Moreover, in 3 cases, 2–3 distinct isolates of a virus species co-infected the same plant.	Polyadenylated RNA	Illumina Genome Analyzer IIx	[47]

Table 3. Cont.

Host	Study finding/virome	Sample preparation/target	Sequencing platform	Ref.
<i>Nicotiana benthamiana</i> , <i>Laodelphax striatellus</i> (small brown leafhopper), rice	The presence <i>Rice stripe virus</i> (RSV) siRNAs was demonstrated in infected rice as well as in infected <i>N. benthamiana</i> and viruliferous <i>L. striatellus</i> . Also, results indicate the potential existence of RNAi-mediated immunity against RSV infection in <i>L. striatellus</i> , a member of <i>Hemipteran</i> that transmits about 55% of the known plant viruses. Moreover, demonstrated that siRNA are generated differentially in different hosts.	siRNAs	Illumina	[48]
Sweet potato	Detected <i>Sweet potato feathery mottle virus</i> strain RC, <i>Sweet potato virus C</i> ( <i>Potyvirus</i> ), <i>Sweet potato chlorotic stunt virus</i> strain WA ( <i>Crinivirus</i> ), <i>Sweet potato leaf curl Georgia virus</i> ( <i>Begomovirus</i> ), and <i>Sweet potato pakakuy virus</i> strain B (synonym: <i>Sweet potato badnovirus</i> B) infecting sweet potato crops in Central America. Also, 4 viruses were detected in a sweet potato sample from the Galapagos Islands. Results suggest that siRNAs deep sequencing analysis is suitable for use as a reliable method for detection of plant viruses in infected crops.	si RNA	Illumina Genome Analyzer	[49]
Pepper (genotype Yolo Wonder)	Developing a mathematical model that estimates genetic drift and selection intensities using next generation sequencing data of 4 variants of <i>Potato virus Y</i> that differ by 1–2 substitutions involved in pathogenicity.	Total RNA	Roche 454	[50]
<i>Arabidopsis thaliana</i>	Analysis of viral siRNAs from DNA virus-infected cells showed that the entire circular genomes of <i>Cauliflower mosaic virus</i> (genus: <i>Caulimovirus</i> , family: <i>Caulimoviridae</i> ) and <i>cabbage leaf curl virus</i> (genus: <i>Begomovirus</i> , family: <i>Geminiviridae</i> ) are densely covered with siRNAs in both sense and antisense polarities without gaps. This would enable <i>de novo</i> reconstruction of the complete DNA virus genomes from siRNAs.	siRNAs	Illumina Genome Analyzer	[51–53]
Black pepper	The complete genome sequence of <i>Piper yellow mosaic virus</i> was determined. It was also established that the virus is a member of the genus <i>Badnovirus</i> and the family <i>Caulimoviridae</i> . Fragments of two additional novel viruses belong to <i>Caulimoviridae</i> were sequenced and the viruses were tentatively named Piper DNA virus 1 and 2.	Viral and plant DNA were isolated from virus-enriched fraction	Roche 454 GS-FLX Titanium	[54]

**Table 4.** Next-generation sequencing of plant viral siRNA, total RNA or dsRNA from virus-infected temperate fruit crop, citrus or fig hosts.

Host	Study finding/virome	Sample preparation/target	Sequencing platform	Ref.
Raspberry	A novel virus isolated from infected raspberry plants was completely sequenced and characterized. It was designated as <i>Raspberry latent virus</i> . The virus is a novel dicot-infecting reovirus in the family <i>Reoviridae</i> , subfamily <i>Spinareovirinae</i> .	dsRNA	Illumina	[55]
Citrus	In <i>Citrus tristeza virus</i> (CTV) infected citrus plants It was shown that the citrus homologues of Dicer-like ribonucleases mediate the genesis of the 21 and 22 nt CTV siRNAs and that the ribonucleases act not only on the genomic RNA but also on the 30 co-terminal subgenomic RNAs and, particularly, on their dsRNA forms. A novel citrus miRNAs was also indentified and how CTV influences their accumulation was determined.	CTV sRNAs, gRNA sgRNAs	Illumina	[56]
Citrus	Genomic organization and other molecular characterizations were determined for <i>Citrus yellow vein clearing virus</i> . Analyses suggested that the virus is the causal agent of yellow vein clearing disease of lemon trees and represent a new species in the genus <i>Mandarivirus</i> .	siRNAs	Illumina	[57]
Citrus	A novel DNA virus species, member of the family <i>Geminiviridae</i> , was identified and associated with citrus chlorotic dwarf disease. A provisional name of Citrus chlorotic dwarf-associated virus was proposed.	siRNAs and total DNA	Illumina HiSeq2000	[58]
Apple, Citrus, Grapevine	Detected ASPV, ACLSV and an unknown mycovirus. Detected two variants of CTV and ASGV. Detected variants of GLRaV-3, GVA and an unknown mycovirus.	siRNAs	Illumina	[59]
Apple	Identified agents associated with green crinkle disease of apple trees. The disease is a complex one as the following viruses were identified associated with it: ASGV, ASPV, ACLSV, ApLV, ApPCLSV and PCMV.	siRNAs	Illumina HiSeq2000	[60]
<i>Prunus</i>	Detected and identified known <i>Prunus</i> viruses such as PPV, PNRS, etc. and novel virus agents.	dsRNA	Roche 454	[61]
Fig	Detected <i>Fig mosaic virus</i> and <i>Fig latent virus-1</i> for their elimination from infected clones. It is the first application of next-generation sequencing technology to detect and identify known and new species of viruses infecting fig trees.	dsRNAs	Illumina	[62]
Blackberry	A novel <i>Ampelovirus</i> in the family <i>Closteroviridae</i> was identified as one of the viruses associated with blackberry yellow vein disease complex.	dsRNAs	Illumina	[63]
Cherry	Characterization of the genome of the divergent <i>Little cherry virus 1</i> (LChV1) isolate and establishing that LChV1 isolates could be responsible for Shirofugen stunt disease syndrome.	dsRNAs	Roche 454 Pyrosequencing multiplex approach	[64]

Table 4. Cont.

Host	Study finding/virome	Sample preparation/target	Sequencing platform	Ref.
Citrus	The complete nucleotide sequence and structure of a novel virus of the genus <i>Cilevirus</i> was determined. The novel virus causes symptoms similar to citrus leprosis and it is suggested to be called Citrus leprosis virus cytoplasmic type 2.	siRNAs	Illumina	[65]
Citrus	A novel virus was discovered by analysis of the contigs assembled from the virus siRNAs sequences which showed similarity with luteovirus sequence, particularly with <i>Pea enation mosaic virus</i> , the type member of the genus <i>Enarnovirus</i> . The complete genome of the virus was determined and the new virus was provisionally named Citrus vein enation virus.	siRNAs	Solexa-Illumina	[66]

Virus abbreviations: ACLSV, *Apple chlorotic leaf spot virus*; ApLV, *Apricot latent virus*; ApPCLSV, *Apricot pseudo-chlorotic leaf spot virus*; ASGV, *Apple stem grooving virus*; ASPV, *Apple stem pitting virus*; CTV, *Citrus tristeza virus*; GLRaV-3, *Grapevine leaf roll associated virus 3*; GVA, *Grapevine virus A*; PCMV, *Peach chlorotic mottle virus*; PNRS, *Prunus necrotic ring spot virus*; PPV, *Plum pox virus*.

Table 5. Next-generation sequencing of plant viral siRNA, total RNA or dsRNA from virus-infected grapevine.

Host	Study finding/virome	Sample preparation/target	Sequencing platform	Ref.
Grapevine	A novel <i>Marafivirus</i> ( <i>Grapevine Syrah 1virus</i> ) was identified associated with grapevine syrah decline. The virus was also identified in leafhopper vector. Also detected in plant tissue GRSPaV, GRVfV, GLRaV-9, and viroids.	Total RNA or dsRNA	Roche 454	[67]
Grapevine	<i>Grapevine virus E</i> , not previously in South Africa. A mycovirus similar to <i>Penicillium chrysogenum virus</i> , two other mycoviruses, GLRaV-3, GRSPaV, GVA.	dsRNA	Illumina	[68]
Grapevine	Viruses of the genera <i>Foveavirus</i> , <i>Maculavirus</i> , <i>Marafivirus</i> , and <i>Nepovirus</i> were detected. siRNAs originate from both genomic and antigenomic strands with the exception of tymoviruses, the majority are derived from antigenic virus strand.	siRNAs	Illumina	[69]
Grapevine	A novel DNA virus was discovered associated with the grapevine vein-clearing and vine decline syndrome. The virus belongs to genus <i>Badnavirus</i> in the family <i>Caulimoviridae</i> . It is the first DNA virus discovered in grapevine. It has been provisionally named Grapevine vein clearing virus.	siRNAs	Illumina Genome Analyzer	[70]
Grapevine	Twenty six fungal groups were identified in a single plant source. Three of the mycoviruses were associated with <i>Botrytis cinerea</i> . Most of the rest were undescribed.	dsRNA	Roche 454	[71]

Table 5. Cont.

Host	Study finding/virome	Sample preparation/target	Sequencing platform	Ref.
Grapevine	A novel species of virus was discovered for which the provisional name Grapevine Pinot gris virus is proposed. Also, detected and identified GRSPaV, GRVFV, GSy 1V, and viroids.	siRNAs	Illumina	[72]
Grapevine	A novel circular DNA virus was identified associated with red blotch disease in grapevine in California. A provisional name of Grapevine red blotch-associated virus is proposed for the novel virus.	dsRNA extracted without DNase treatment	Illumina Genome Analyzer IIx	[73]
Grapevine	A novel <i>Vitivirus</i> was identified. The virus is provisionally named Grapevine virus F.	dsRNA	Illumina Genome Analyzer IIx	[74]
Grapevine	Characterization of siRNAs associated with grapevine leafroll disease.	siRNAs	Illumina	[75]
Grapevine	Complete sequence of a novel single-stranded DNA virus associated with grapevine red leaf disease (GRD). The virus is tentatively named Grapevine red leaf-associated virus (GRLaV). The virus represents an evolutionary distinct lineage in the family <i>Geminiviridae</i> . Also detected in plant tissue GRSPaV, GFV, and viroids.	Total RNA treated with DNase	Illumina Genome Analyzer IIx	[76]

Virus abbreviations: GFV, *Grapevine fanleaf virus*; GLRaV-3, *Grapevine leaf roll-associated virus 3*; GLRaV-9, *Grapevine leaf roll-associated virus 9*; GRSPaV, *Grapevine rupestris stem pitting-associated virus*; GRVFV, *Grapevine rupestris vein feathering virus*; GSy 1V, *Grapevine Syrah 1 virus*; GVA, *Grapevine virus A*.

**Table 6.** Next-generation sequencing of isolated siRNA or dsRNA from viroid-infected grapevine, peach, cucumber, *Nicotiana benthamiana* or tomato hosts.

Host	Study finding/virome	Sample preparation/target	Sequencing platform	Ref.
Grapevine	Different Dicer-like enzymes target RNAs of <i>Hop stunt viroid</i> , <i>Grapevine yellow speckle viroid 1</i> . Also, study suggested that the viroid RNAs may interact with host enzymes involved in the RNA-directed DNA methylation pathway	siRNAs	Solexa, Illumina	[77]
Grapevine	Detection and identification of <i>Australian grapevine viroid</i> , <i>Hop stunt viroid</i> and <i>Grapevine yellow speckle viroid</i>	Total RNA or dsRNA	Roche 454	[67]
Peach	To study the genesis of <i>Peach latent mosaic viroid</i> siRNA and viroid pathogenesis	siRNAs	Illumina	[78]

Table 6. Cont.

Host	Study finding/virome	Sample preparation/target	Sequencing platform	Ref.
<i>Nicotiana benthamiana</i>	RNA-dependent RNA polymerase 6 restricts accumulation and precludes meristem invasion of <i>Potato spindle tuber viroid</i> which replicates in nuclei	Plant and viroid siRNAs	Illumina EAS269 GAII	[79]
Cucumber	To study <i>Hop stunt viroid</i> pathway involved in the biogenesis of the viroid siRNAs	siRNAs	Illumina	[80]
Tomato	Detection and identification of <i>Potato spindle tuber viroid</i>	siRNAs	Illumina Genome Analyzer IIx	[44]
Grapevine	Detection and identification of <i>Grapevine yellow speckle viroid 1</i> and <i>Hop stunt viroid</i>	siRNAs	Illumina	[72]
Grapevine	Detection and identification of <i>Grapevine yellow speckle viroid 1</i> and <i>Hop stunt viroid</i>	siRNAs and dsRNAs	Illumina	[81]
Grapevine	Discovery of viroid-like circular RNA 375 nt long with hammerhead ribozymes. Currently, infectivity studies showed that the RNA is not infectious which may suggest that it is viral satellite	siRNAs	Illumina	[82]
Grapevine	Characterization of siRNAs of <i>Hop stunt viroid</i> , <i>Grapevine yellow speckle viroid 1</i> , and <i>Grapevine yellow speckle viroid 2</i>	siRNAs	Illumina	[75]
Grapevine	Detection and identification of <i>Grapevine yellow speckle viroid 1</i> , <i>Hop stunt viroid</i> , <i>Citrus exocortis Yucatan viroid</i> and <i>Citrus exocortis viroid</i> from both symptomatic and non-symptomatic samples of grapevine read leaf disease	Total RNA treated with Dnase	Illumina Genome Analyzer IIx	[76]

Table 7. Next-generation sequencing of RNA viruses, DNA viruses or bacteria DNA from insect vectors.

Vector	Study finding/virome	Sample preparation/target	Sequencing platform	Ref.
Grapevine leafhopper	A novel <i>Merafivirus</i> associated with grapevine syrah decline was detected in the vector	Total nucleic acids of the viruliferous vector	Roche 454	[67]
Citrus psyllid	A complete genomic sequence of the bacterium, “Candidatus Liberibacter asiaticus” was obtained. The genome is circular and its size is about 1.23 Mb, The bacterium is the causal agent of citrus Huanglongbing (greening) disease	DNA extracted from a single “Ca. L. asiaticus”-infected Asian citrus psyllid ( <i>Diaphorina citri</i> )	Roche 454 GS-FLX	[83]
<i>Bemisia tabaci</i>	Four novel <i>Begomovirus</i> species were discovered in their viruliferous vectors	Purified viral DNA	Metagenomic reads 100–700 nt	[84]
<i>Aodelphgax striatellus</i> (small brown leafhopper)	The presence of <i>Rice stripe virus</i> siRNAs was demonstrated in the viruliferous vector <i>L. striatellus</i>	siRNAs	Illumina	[48]

**Table 8.** Next-generation sequencing of isolated DNA, RNA or miRNA from grapevine, periwinkle or citrus infected with phytoplasmas or difficult to culture bacteria.

Host	Study finding/pathogen	Sample preparation/target	Sequencing platform	Ref.
Grapevine	It was demonstrated that sequences of infected phytoplasmas belonged to 16 SrV and 16 SrXII groups, as well as to <i>Candidatus</i> Phytoplasma prunorum (16SrX-B) whereas some sequences could not be assigned to a single phytoplasma group. Also a high number of single nucleotide polymorphisms (SNPs) were found. Suggested NGS may be used for future phytoplasma detection in quarantine.	Total DNA from mid-vein leaf tissue	Amplicon sequencing by Roche 454 GS FLX	[85,86]
Grapevine	Demonstrated significant changes in the transcriptome of Aster yellows phytoplasma-infected Grapevine cv. Chardonnay. The study could contribute to understanding the unknown mechanisms of phytoplasma pathogenicity.	Total RNA and DNA	Illumina HiSeq 2000	[87]
Periwinkle	Genomic analysis of four phytoplasma strains of 16SrIII group and two strains of the 16SrI-B subgroup revealed the significant role of horizontal gene transfer among different “Ca. Phytoplasma” species in shaping phytoplasma genomes and promoting their diversity.	Standard DNA preparation from infected periwinkle	Illumina	[88]
Citrus (Mexican lime)	Identified miRNA families that are expressed differentially upon infection of Mexican lime trees with <i>Candidatus</i> Phytoplasma aurantifolia. The study increases our understanding of the molecular basis of witches’ broom disease which may lead to development of new strategies for its control.	miRNAs were isolated from infected and from healthy tissues	Illumina HiSeq 2000	[89]
Citrus	Demonstrated that several miRNAs and siRNAs were highly induced by <i>Ca. L. Asiaticus</i> (Las) infection, which can be potentially developed into early diagnosis markers of huanglongbing (HLB) disease (citrus greening disease). MiR399 was induced specifically by infection of Las. MiR399 is induced by phosphorous starvation in other plant species. Applying phosphorous significantly reduced HLB symptoms in citrus.	RNA fragments 18–28 nt in length obtained after fractionation of total RNA on denaturing polyacrylamide gel electrophoresis. The total RNA was extracted from Las-infected tissue with HLB symptoms	Illumina	[90]

## 7. Bioinformatics' Software Tools for Data Analysis

The software functions for next-generation sequencing data analysis may be classified into four general categories which include alignment of sequence reads, base-calling and/or polymorphism detection, *de novo*, and genome browsing and annotation. Several software packages have been developed for each category. For example, a variety of bioinformatics' software tools for short-read sequencing used for *de novo* assembly of genomes and transcriptomes are available [91,92]. A practical application of *de novo* genome assembly software tools has been also recently reported [93]. New categories and new software tools for short-read sequencing analysis are continuously being developed commercially world-wide, especially in the U.S., Europe and Australia. Many commercial sequencing companies, *i.e.*, Roche, Illumina, SOLiD, *etc.*, offer their software services to customers for different types of DNA or RNA sequencing analysis. It is worth mentioning that alignment software such as BLAST or BLAT are not suitable for short-read sequencing analysis of next generation sequencing as they were developed specifically for long reads generated by conventional first generation sequencing.

A full review of software tools available for quality control assembly and quantitative analysis of next-generation sequencing is beyond the scope of this article, but have been the subject of several review articles, *i.e.*, [94–97] and a book [98]. Moreover, the journal *Bioinformatics* devoted one of its volumes published in 2009 (volume 25, No.4) to bioinformatics tools and algorithms that have been developed for next-generation sequencing data analysis. These bioinformatics software tools and algorithms are continuously being developed and improved to keep abreast and pace with the latest advances of next-generation sequencing technologies.

## 8. Cost of DNA Sequencing

According to the latest information released by the National Human Genome Research Institute, U.S. National Institutes of Health, the cost per raw megabase of DNA sequence and the cost per genome were reduced dramatically from July 2001 to July 2013 [99]. During this time period, first generation sequencing methods were used from 2001 through 2007, then, second generation NGS platforms were used from 2008 to present, 2013. The cost per raw megabase of DNA sequence was reduced from about \$8,000 in 2001 to \$700 in 2007 and then to less than \$0.1 in 2013 (Figure 1, top). Similarly, the cost per genome DNA sequence went from \$100, 000,000 in 2001 to \$10,000,000 in 2007 then to about \$8,000 in 2013 (Figure 1, bottom).

In the next few years, it is expected that the third generation of DNA sequencing platforms will increase sequencing capacity and speed while reducing cost. A genome may be sequenced for about \$1,000 or less in about 15 minutes. That is 85% less than the current cost.

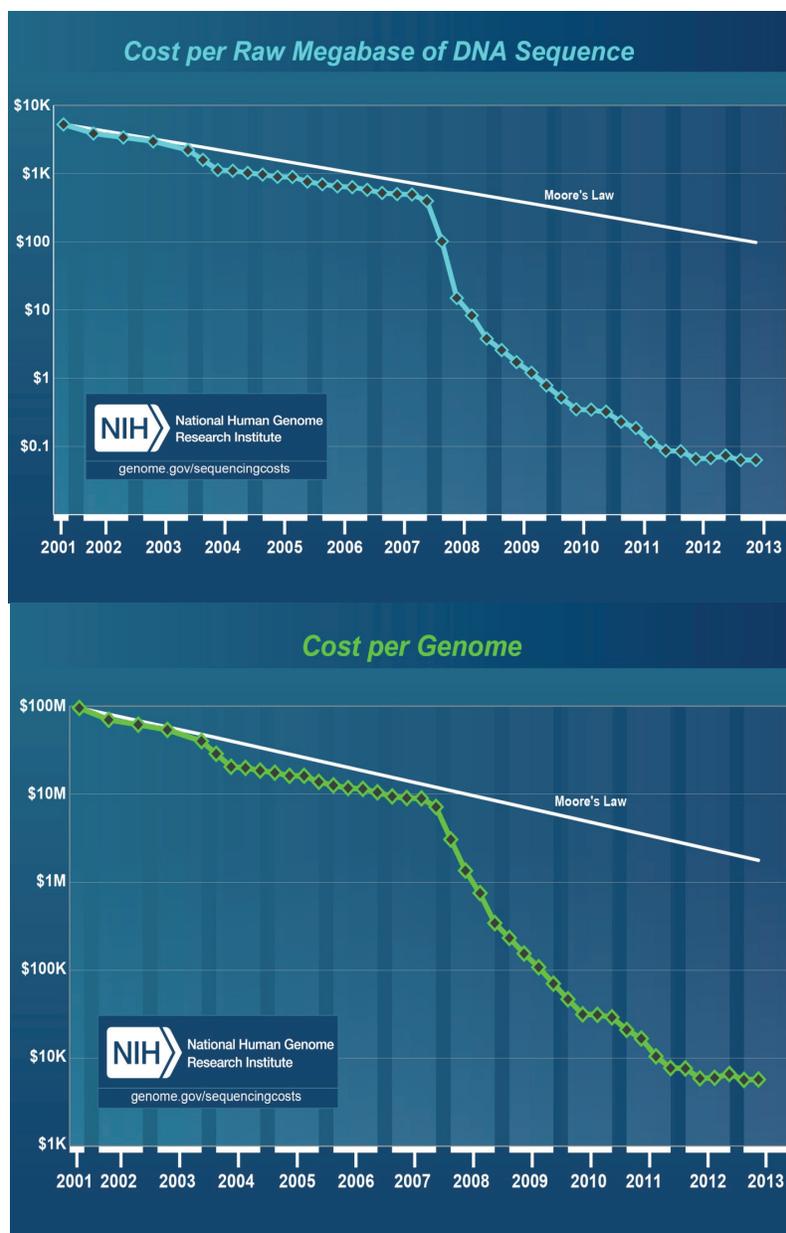
## 9. Biological Applications of Next-Generation Sequencing

To date the biological applications of next-generation sequencing focused primarily on:

Full or complete genome sequencing (*de novo* sequencing and/or resequencing genomics): Its aim is to sequence the entire genome of an organism such as humans, primates, dogs, cats, mice, nematodes, fungi, bacteria, viruses, *etc.* It is worth noting that unlike full genome sequencing, DNA

profiling only determines the likelihood that genetic material came from a particular individual or group; single nucleotide polymorphism (SNP) genotyping covers less than 0.1% of the genome.

**Figure 1.** Cost in US Dollars per raw megabase of DNA sequence (**top**) and cost per genome (**bottom**) from July 2001 to July 2013 as estimated by the National Human Genome Research Institute, U.S. National Institutes of Health, Bethesda, MD, USA. During this time period, first generation sequencing methods were used from 2001 through 2007, then NGS platforms were used from 2008 to present.



Metagenomics: The term “metagenomics” was first used in publications in 1998. Consequently in 2005 Kevin Chen and Lior Pachter at University of California (Berkeley, CA, USA), defined metagenomics as “the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, by passing the need for isolation and lab

cultivation of individual species.” The goals of this approach are to characterize organisms present in a sample and identify what roles each organism has within a specific natural environment. Metagenomics microbial samples are found almost everywhere such as the human body, soil samples, plant samples, water samples, *etc.*

**Transcriptome sequencing:** It includes sequencing and analysis of full-length mRNA and small RNA (microRNA profiling and discovery); mRNA transcript expression analysis (full-length mRNA, expressed sequence tags [ESTs] and ditags, and allele-specific expression); quantification of gene expression and alternative splicing, transcript annotation, discovery of transcribed SNPs or somatic mutation.

Sequencing of the cDNAs of diverse RNAs (genomic, mRNA, micro RNA, ribosomal RNA, *etc.*) has been coined RNA-seq. [100].

**Amplicon sequencing:** This method allows the detection of mutations at extremely low frequency levels in PCR amplified specific targeted regions of DNA. It has been used to identify low frequency somatic mutations in cancer samples, discovery of rare variants in HIV-infected patients, *etc.* It also has been used recently for sequencing phytoplasmas from grapevines.

Other applications of NGS in specific categories may include reduced representation sequencing, (*i.e.*, large-scale polymorphism discovery), targeted genomic resequencing, (*i.e.*, targeted polymorphism and mutation discovery), paired end sequencing, (*i.e.*, discovery of inherited and acquired structural variation), sequencing of bisulfite-treated DNA, (*i.e.*, for determining patterns of cytosine methylation in genomic DNA), chromatin immunoprecipitation sequencing (ChIP-Seq), (*i.e.*, genome-wide mapping of protein-DNA interactions), nuclease fragmentation and sequencing, (*i.e.*, nucleosome positioning) [92]. In addition, other applications include: pharmogenomics sequencing to determine suitable drugs for individuals; diagnosis and discovery of new human diseases; cancer therapy; humans origin, migration of humans, evolution, and others.

## 10. Applications of Next-Generation Sequencing in Medical Virology

The first applications of next-generation sequencing in the field of medical virology were to monitor population diversity in HIV [101,102]. In September 2008, an outbreak of unexplained hemorrhagic fever in people was reported in South Africa which resulted in death of four out of five patients. Such outbreaks required a rapid response to control both infection and public anxiety. NGS of infected blood serum samples revealed the discovery of a novel *arenavirus* [103]. Rapid pathogen discovery using NGS is now available to monitor all outbreaks, regardless of the disease.

Currently, next-generation sequence technologies applications in medical virology may include, but are not limited to, determination of: full-length viral genome sequence; study of viral genome variability; characterization of viral quasi-species; viral evolution. Metagenomics-based strategies have been used for detection of unknown viral pathogens, discovery of novel viruses, detection and identification of tumor viruses; characterization of the human “viral metagenome” or “virome” in healthy and disease conditions and epidemiology of viral infection. These strategies have also been used for analysis of resistance profiles to drugs and host immunity; quality control of viral vaccines; transcriptomics studies for measurements of mRNA to achieve new insights into genome expression

and how this may be modified in health and disease as well as for exploring the role of microRNAs in virus replication and pathogenesis (for review, see [21,104]).

## 11. Current Applications of Next-Generation Sequencing in Plant Virology

NGS technologies combined with sophisticated bioinformatics have been recently changing the field of plant virology, particularly in the areas of genome sequencing, ecology, discovery, epidemiology, transcriptomics, replication, detection and identification. Currently a small number of plant RNA viruses and viroids have been identified from infected tissues and sequenced by RNA-seq. Either total nucleic acid or total double-stranded RNA (dsRNA) from pathogen infected tissue was isolated and the virus or viroid was identified by NGS; alternatively, in some cases, host nucleic acid was partially eliminated by hybridization to nucleic acid isolated from healthy plants to enrich for virus sequences in the infected plant material prior to sequencing. All the viruses and/or viroids detected and identified in infected plants or viruses in viruliferous vectors using NGS are termed “virome”.

Plant viruses or viroids can also be detected indirectly. In response to infection by RNA/DNA viruses or viroids, the host plant generates specific RNA molecules, 21 to 24 nt in length, called short interfering RNAs (siRNA). RNA silencing (RNAi) is a cytoplasmic cell surveillance system to recognize dsRNA and specifically destroy single and double stranded RNA molecules homologous to the inducer, using small interfering RNAs as a guide (for review see [105,106]). NGS of siRNAs offers good opportunities to identify viruses or viroids infecting plants, even at extremely low titers, in symptomless infections, and including previously unknown viruses or viroids. Next-generation sequencing can provide thousands to millions of siRNA sequences from virus or viroid infected plant materials. When virus or viroid-derived siRNAs are abundant enough, virus or viroid genome fragments can be assembled. Since the virus or viroid siRNAs are 21–24 nt in length, their sequences can be employed directly as primer sequences to amplify viral or viroid fragments by PCR or RT-PCR.

Tables 3–6 summarize the use of next generation technologies in analyzing virus or viroid virome in infected plant species. NGS technologies have also been used for detection and identification of virus virome in viruliferous vectors (Table 7) and insect viruses [107], and the bacterial “virome” of the bacterium ‘*Candidatus Liberibacter asiaticus*’ in its insect vector (Table 7) and infected host plant (Table 8). In addition, these technologies have been used for analyzing phytoplasma “virome” in infected plants (Table 8).

## 12. Next-Generation Sequencing and Revealing the Etiology of Unknown Diseases and Latent Infections

The recent use of NGS technologies in plant virology revealed that some diseases of unknown etiology that affect herbaceous and grass hosts or latent agent infections in wild hosts of different species are caused by new or known viruses (Table 3). This table shows that the following novel viruses were discovered: (a) in sweet potato, two dsDNA viruses (*Badnavirus*) and one ssDNA virus (*Mastrevirus*); (b) in *Gomphrena globosa* mechanically inoculated with sap from *Liatris spicata*, a *Cucumovirus* provisionally named Gayfeather mild mottle virus; (c) in wild plant species from 15 families, unclassified virus families and numerous novel virus species; (d) in wild cocksfoot grass, *Cereal dwarf*

virus (*Luteovirus*) discovered infecting this plant species; (e) in pepper and eggplant, *Pepper yellow leaf curl virus* (*Polerovirus*) and *Eggplant mild leaf mottle* (*Ipomovirus*), respectively; (f) watercress, a virus provisionally named Watercress white vein virus; (g) in tomato, a *Potyvirus* species with the proposed name Tomato necrotic stunt virus, also in tomato in a different study, an unspecified *Tospovirus* and a squash-infecting geminivirus; (h) in 17 wild plant species among them 14 Australian indigenous species, four novel viruses (*Potyvirus*, *Sadwavirus*, *Trichovirus*); (i) in black pepper, two DNA viruses (*Caulimoviridae*) with proposed names DNA virus 1 and DNA virus 2.

Most, if not all, diseases of unknown etiology that affect stone, pome and small fruit crops, as well as citrus, and grapevine are graft transmissible. That may imply that systemic pathogens are most likely involved in the etiology of these diseases. Currently there are over 35 diseases of unknown etiology that affect stone fruits and a similar number that affect pome fruits [108]. As revealed recently by next generation sequencing, the disease apple green crinkle is a disease complex and at least 6 known viruses—the 3 apple latent viruses and 3 stone fruit viruses—are associated with it (Table 4). Also, it was established that Shirofugen stunt disease syndrome could be caused by *Little cherry virus 1* (Table 4). Among major diseases of unknown etiology that affect small fruits are blueberry fruit drop, blueberry bronze leaf curl, blackberry yellow vein, raspberry leaf curl, raspberry yellow spot and strawberry latent C [109]. Quito-Avila *et al.* [55] discovered a novel reovirus in the family *Reoviridae*, subfamily *Spinareovirinae* infecting raspberry; Thekke-Veetil *et al.* [63] identified a novel *Ampelovirus* as one of the viruses associated with blackberry yellow vein disease (Table 4).

In citrus, a novel DNA geminivirus was discovered and provisionally named Citrus chlorotic dwarf-associated virus; two novel RNA viruses were also discovered, one was provisionally named Citrus leprosis virus cytoplasmic type 2 (*Cilevirus*) and the other was named Citrus vein enation virus (*Enamovirus*) (Table 4).

NGS has also been used for detection and identification of virus virome in viruliferous vectors as well as of the bacterium, “*Candidatus Liberibacter asiaticus*” in its vector (Table 7). It has also been utilized for the detection and discovery of insect viruses [107].

In addition to viruses and viroids, NGS technologies have been used for analyzing phytoplama “virome” in infected grapevine, periwinkle or citrus (Table 8). Total DNA and/or RNA or siRNAs from infected host were used in these studies. Moreover, it was also demonstrated that, upon citrus infection by the hard to grow bacterium, “*Candidatus Liberibacter asiaticus*”, the causal agent of citrus Huanglongbing (greening) disease, several miRNAs and siRNAs were highly induced and can be potentially developed into early diagnosis markers of the disease (Table 8).

Next generation sequencing has been shown to be a good option for investigating diseases of unknown etiology in grapevine [81]. A novel *Marafivirus* (*Grapevine Syrah 1 virus*) was identified associated with grapevine syrah decline (Table 5). In addition, the following novel RNA viruses were discovered infecting grapevine and given the provisional names: Grapevine Pinto gris virus, Grapevine virus F. (*Vitivirus*). Two DNA viruses were also discovered, the first belongs to the genus *Badnavirus*, family *Caulimoviridae*, provisionally named Grapevine vein clearing virus; the second is a DNA geminivirus and given the provisional name Grapevine red-blotch-associated virus or Grapevine red-leaf-associated virus (Table 5). A viroid-like circular RNA with a hammerhead structure was discovered in infected grapevine (Table 6). Currently, infectivity studies showed that the RNA is not infectious which may suggest it is a viral satellite.

### 13. Prospective

NGS technologies have become available during the last several years and have been widely used in whole genome sequencing, metagenomics, RNA sequencing, *etc.* The technologies are in continuous development and improvement. Sequencing chemistries are becoming established, maturing and evolving, read lengths and fidelity are increasing, allowing us to investigate thoroughly viral genome diversity, virus-host interaction, virus epidemiology, virus diagnosis and elimination, *etc.* The major sequencing platforms are being developed to be easier to use and more cost effective. Moreover, very recently Roche, Illumina, Life Technologies Ion Torrent and Oxford Technologies Nanopore sequencing companies have developed and released in the market reasonably priced bench top platforms with the aim of making next-generation sequencing technologies more readily available to more research laboratories as well as to diagnostic laboratories for viral, bacterial and fungal human, animal, and plant diseases. In 2012–2013 single-molecule DNA sequencing in a miniaturized disposable device for single use has been developed by Oxford Technologies Nanopore that allows simple and cheap high throughput sequencing. It is worth mentioning that Nanopore sequencing platform models have the potential to rapidly generate ultra long single molecule reads.

The most recent report of the International Committee for the Taxonomy of Viruses lists about 900 plant viruses [110]. The utilization of NGS in plant virology will definitely increase this number very significantly as new viruses are being discovered and characterized in different plant host species, including wild ones, as well as in different insect vectors. Conventional serological or molecular detection and identification methods of plant viruses or viroids depend on prior knowledge of antibody or sequence of the virus or viroid of interest. The NGS technologies have provided a very powerful alternative for detection and identification of these pathogens without a *priori* knowledge. Metagenomics developed by NGS technologies has been proven to be sensitive, accurate and fast in detection and identification of known and unknown viral and viroid viromes without bias in infected plant species, including woody perennial crops which have low titers of these pathogens. Thus, NGS technologies have the potential to be used in quarantine and certification programs of grapevine, pome and stone fruits, small fruits as well as citrus. The potential use of these technologies in diagnostics has been recently suggested in temperate fruit crops [111], citrus [112], grapevine and other crops [113]. Data generated by these technologies can be used effectively to improve efficiency and reliability of these programs as well as in programs aimed at virus and viroid elimination from vegetatively propagative material. Accordingly, NGS technologies will be a significant and powerful tool in controlling virus and viroid diseases.

### 14. Conclusions

The next-generation high throughput sequencing technologies have been available for several years. These technologies and bioinformatics provide rapid and low cost DNA sequencing for biological material, including plant viruses and viroids. NGS platforms, which have different underlying biochemistries and differ in sequencing protocol, produce large amounts (typically millions to billions) of generally short DNA sequence reads, length typically between 25 bp and 400 bp. These reads are shorter than the 750 bp first-generation Sanger sequences. The beginning of the era of NGS technologies was in 2000 and the first marketed sequencing platform was in 2004. Since then the NGS

industry has expanded with several companies marketing different models of sequencing platforms, most recently the bench top platforms. The sequencing platforms are continually improved to become faster, more efficient and cheaper in order to bring next-sequencing technologies into many more laboratories to expand their use in biological research and diagnostics.

NGS technologies combined with sophisticated bioinformatics were successfully utilized in plant virology since 2009, particularly in the areas of genome sequencing, detection and identification, discovery, transcriptomics, replication, ecology, and epidemiology. Known and novel plant RNA and DNA viruses and their satellites as well as viroids and phytoplasmas from different infected plant species were successfully studied using these technologies. The number of published studies, however, are still small but significant (Tables 3–8). The utilization of NGS in plant virology in the near future will definitely increase whether in research or diagnostics. The potential use of these technologies in plant certification and quarantine programs can effectively improve the efficiency and reliability of these programs and in controlling virus and viroid diseases at both the national and international levels.

### Acknowledgments

The authors would like to acknowledge our colleagues Rosemary W. Hammond and Moshe Bar-Joseph for reviewing the manuscript.

### Conflicts of Interest

The authors declare no conflict of interest

### References and Notes

1. Watson, J.; Crick, F.H.C. A structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737–738.
2. Holley, R.W.; Madison, J.T.; Zamir, A. A new method for sequence determination of large oligonucleotides. *Biochem. Biophys. Res. Commun.* **1964**, *17*, 389–394.
3. Holley, R.W.; Apgar, J.; Everett, G.A.; Madison, J.T.; Marquisee, M.; Merrill, S.H.; Penswick, J.R.; Zamir, A. Structure of a ribonucleic acid. *Science* **1965**, *147*, 1462–1465.
4. Sanger, F.; Air, G.M.; Barrell, B.G.; Brown, N.L.; Coulson, A.R.; Fiddes, J.C.; Hutchison, C.A.; Slocombe, P.M.; Smith, M. Nucleotide sequence of bacteriophage phX174DNA. *Nature* **1977**, *265*, 687–695.
5. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminator inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467.
6. Maxam, A.M.; Gilbert, W.A. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 560–564.
7. Jackson, D.; Symons, R.H.; Berg, P. Biochemical method for inserting new genetic information into DNA of Simian virus 40: Circular SV40 DNA molecules containing Lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 2904–2909.
8. Gilbert, W.; Maxam, A. The nucleotide sequence of the lac operator. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3581–3584.

9. Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; Hatfull, G.; Hudson, G.S.; Satchwell, S.C.; Seguin, C.; *et al.* DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* **1984**, *310*, 207–211.
10. Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.; Kerlavage, A.R.; Bult, C.J.; Tomb, J-F.; Dougherty, B.A.; Merrick, J.M.; *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **1995**, *269*, 496–512.
11. Ronaghi, M.; Karamohamed, S.; Pettersson, B.; Uhlén, M.; Nyrén, P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **1996**, *242*, 84–89.
12. Kawashima, E.; Farinelli, L.; Mayer, P. Method of Nucleic Acid Amplification. Int. Appl. No. PCT/GB98/00961, 8 October 1998.
13. Ewing, B.; Green, P. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genomic Res.* **1998**, *8*, 186–194.
14. Life Sciences, a Roche Company. Available online: <http://www.454.com/> (accessed on 22 October 2012).
15. Illumina (Illumina). Available online: [http://www.illumina.com/technology/sequencing\\_technology/ilmn/](http://www.illumina.com/technology/sequencing_technology/ilmn/) (accessed on 22 October 2012).
16. SOLiD (Life Technologies). Available online: <http://www.lifetechnologies.com/us/en/home/life-science/sequencing/next-generation-sequencing/solid-next-generation-sequencing.html/> (accessed on 24 October 2012).
17. Ion-Torrent (Life Technologies). Available online: <http://www.lifetechnologies.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-orkflow/ion-torrent-next-generation-sequencing-run-sequence.html/> (accessed on 25 October 2012).
18. PacBio (Pacific Bioscience). Available online: <http://www.pacificbiosciences.com/> (accessed on 3 November 2012).
19. Nanopore (Oxford Technologies). Available online: <https://www.nanoporetech.com/> (accessed on 5 November 2012).
20. Loman, N.J.; Misra, R.V.; Dallman, T.J.; Constantinidou, C.; Gharbia, S.E.; Wain, J.; Pallen, M.J. Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **2012**, *30*, 434–439.
21. Radford, A.D.; Chapman, D.; Dixon, L.; Chantrey, J.; Darby, A.C.; Hall, N. Application of next-generation sequencing technologies in virology. *J. Gen. Virol.* **2012**, *93*, 1853–1868.
22. Liu, L.; Li, Y.; Li, S.; Hu, N.; He, Y.; Pong, R.; Lin, D.; Lu, L.; Law, M. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, *2012*, doi:10.1155/2012/251364.
23. Archer, J.; Baillie, G.; Watson, S.J.; Kellam, P.; Rambaul, A.; Robertson, D.L. Analysis of high-depth sequence data for studying viral diversity: A comparison of next generation sequencing platforms using Segminator II. *BMC Bioinform.* **2012**, *13*, doi:10.1186/1471-2105-13-47.
24. Kirsher, M.; Kelso, J. High-throughput DNA sequencing-concepts and limitations. *Bioassays* **2010**, *32*, 524–536.
25. Carniero, M.O.; Russ, C.; Ross, M.G.; Gabriel, S.B.; Nasbaum, C.; DePristo, M.A. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* **2012**, *13*, doi:10.1186/1471-2164-13-375.

26. Bragg, L.M.; Stone, G.; Butler, M.K.; Hugenholtz, P.; Tyson, G.W. Shining a light on dark sequencing: Characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.* **2013**, *9*, e1003031.
27. Kreuze, J.F.; Perez, A.; Untiveros, M.; Quispe, D.; Fuentes, S.; Barker, I.; Simon, R. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology* **2009**, *388*, 1–7.
28. Adams, I.P.; Glover, R.H.; Monger, W.A.; Mumford, R.; Jackeviciene, E.; Navalinskiene, M.; Samuitiene, M.; Boonham, N. Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. *Mol. Plant. Pathol.* **2009**, *10*, 537–545.
29. Qi, X.; Bao, F.S.; Xie, Z. Small RNA deep sequencing reveals role *Arabidopsis thaliana* for RNA-dependent RNA polymerases in viral siRNA biogenesis. *PLoS One* **2009**, *4*, e4971.
30. Donaire, L.; Wang, Y.; Gonzales-Ibeas, D.; Mayer, K.F.; Aranda, M.A.; Liave, C. Deep sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology* **2009**, *392*, 203–214.
31. Monger, W.A.; Alicai, T.; Ndunguru, J.; Kinyua, Z.M.; Potts, M.; Reeder, R.H.; Miano, D.W.; Adams, I.P.; Boonham, N.; Glover, R.H.; *et al.* The complete genome sequence of the Tanzanian strain of Cassava brown streak virus and comparison with the Ugandan strain sequence. *Arch. Virol.* **2010**, *155*, 429–433.
32. Szittyá, G.; Moxon, S.; Pantaleo, V.; Toth, G.; Rusholme, P.R.L.; Moulton, V.; Burgyan, J.; Dalmay, T. Structural and functional analysis of viral siRNAs. *PLoS Pathog.* **2010**, *6*, e1000838.
33. Roossinck, M.J.; Saha, P.; Wiley, G.B.; Quan, J.; White, J.D.; Lai, H.; Chavarria, F.; Shen, G.; Roe, B.A. Ecogenomics: Using massively parallel pyrosequencing to understand virus ecology. *Mol. Ecol.* **2010**, *19*, 81–88.
34. Pallett, D.W.; Ho, T.; Cooper, I.; Wang, H. Detection of Cereal yellow dwarf virus using small interfering RNAs and enhanced infection rate with Cocksfoot streak virus in wild cocksfoot grass (*Dactylis glomerata*). *J. Virol. Methods* **2010**, *168*, 223–227.
35. Lin, K.-Y.; Cheng, C.-P.; Chang, B.C.-H.; Wang, W.-C.; Huang, Y.-W.; Lee, Y.-S.; Huang, H.-D.; Hsu, Y.-H.; Lin, N.-S. Global analysis of small interfering RNAs derived from Bamboo mosaic virus and its associated satellite RNAs in different plants. *PLoS One* **2010**, *5*, e11928.
36. Yan, F.; Zhang, H.; Adams, M.; Yang, J.; Peng, J.; Antoniw, J.; Zhou, Y.; Chen, J. Characterization of siRNAs derived from rice stripe virus in infected rice plants by deep Sequencing. *Arch. Virol.* **2010**, *155*, 935–940.
37. Silva, T.F.; Romanel, E.A.C.; Andrade, R.R.S.; Farinelli, L.; Osteras, M.; Deluen, C.; Correa, R.L.; Schrago, C.E.G.; Vaslin, M.F.S. Profile of small interfering RNAa from cotton plants infected with the polerovirus Cotton leafroll dwarf virus. *BMC Mol. Biol.* **2011**, *12*, doi:10.1186/1471-2199-12-40.
38. Wylie, S.J.; Jones, M.G.K. The complete genome sequence of a Passion fruit woodiness virus isolate from Australia determined using deep sequencing, and its relationship to other potyviruses. *Arch. Virol.* **2011**, *156*, 479–482.

39. Dombrovsky, A.; Glanz, E.; Sapkota, R.; Lachman, O.; Bronstein, M.; schnitzer, T.; Antignus, Y. Next-generation sequencing a rapid and reliable method to obtain sequence data of the genomes of undescribed plant viruses. In Proceedings of the BARD-Sponsored Workshop—Microarrays and Next-Generation Sequencing for Detection and Identification of Plant Viruses, Beltsville, MD, USA, 17–19 November 2011; Abstract No. 24.
40. Mumford, R.; Adams, I.P.; Glover, R.; Hany, U.; Boonham, N. From high throughput to high output clondiag microarrays and 454 sequence for viral detection and discovery. In Proceedings of the ARD-Sponsored Workshop—Microarrays and Next-Generation Sequencing for Detection and Identification of Plant Viruses, Beltsville, MD, USA, 17–19 November, 2011; Abstract No. 17.
41. Hagen, C.; Frizzi, A.; Kao, J.; Jia, L.; Huang, M.; Zhang, Y.; Huang, S. Using small RNA sequences to diagnose, sequence, and investigate the infectivity characteristics of vegetable-infecting viruses. *Arch. Virol.* **2011**, *156*, 1209–1216.
42. Hu, Q.; Hollunder, J.; Niehl, A.; Korner, C.J.; Gereige, D.; Windels, D.; Arnold, A.; Kuiper, M.; Vasquez, F.F.; Pooggin, M.; *et al.* Specific impact of Tobamavirus infection on the *Arabidopsis* small RNA profile. *PLoS One* **2011**, *6*, e19549.
43. Yang, X.; Wang, Y.; Guo, W.; Xie, Y.; Xie, Q.; Fan, L.; Zhou, X. Characterization of small interfering RNAs derived from the geminivirus/betasatellite complex using deep sequencing. *PLoS One* **2011**, *6*, e16928.
44. Li, R.; Gao, S.; Hernandez, A.G.; Wechter, W.P.; Fei, Z.; Ling, K.-S. Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. *PLoS One* **2012**, *7*, e37127.
45. Lu, J.; Du, Z.-X.; Kong, J.; Chen, L.-N.; Qiu, Y.-H.; Li, G.-F.; Meng, X.-H.; Zhu, S.-F. Transcriptome analysis of *Nicotiana tabacum* infected by Cucumber mosaic virus during systemic symptom development. *PLoS One* **2012**, *7*, e43447.
46. Adams, I.P.; Miano, D.W.; Kinyua, Z.M.; Wangai, A.; Kimani, E.; Phiri, N.; Reeder, R.; Harju, V.; Glover, R.; Hany, U.; *et al.* Use of next generation sequencing for the identification and characterization of *Maize chlorotic mottle virus* and *Sugarcane mosaic virus* causing lethal necrosis in Kenya. *Plant Pathol.* **2012**, doi:10.1111/j.1365-3059.2012.02690x.
47. Wylie, S.J.; Luo, H.; Li, H.; Jones, M.G.K. Multiple polyadenylated RNA viruses detected in pooled cultivated and wild plant samples. *Arch. Virol.* **2012**, *157*, 271–284.
48. Xu, Y.; Huang, L.; Fu, S.; Wu, J.; Zhou, X. Population diversity of Rice stripe virus—Derived siRNAs in three different hosts and RNAi-based antiviral immunity in *Laodelphax striatellus*. *PLoS One* **2012**, *7*, e46238.
49. Kashif, M.; Pietila, S.; Artola, K.; Tugume, A.K.; Makinen, V.; Valkonen, J.P.T. Detection of viruses in sweetpotato from Honduras and Guatemala augmented by deep-sequencing of small-RNAs. *Plant Dis.* **2012**, *96*, 1430–1437.
50. Fabre, F.; Montarry, J.; Coville, J.; Senoussi, R.; Simon, V.; Murry, B. Modelling the evolutionary dynamics of viruses within their hosts: A case study using high-throughput sequencing. *PLoS Pathog.* **2012**, *8*, e1002654.

51. Blevins, T.; Rajeswaran, R.; Aregger, M.; Borah, B.K.; Schepetilnikov, M.; Baerlocher, L.; Farinelli, L.; Meins, F., Jr.; Hohn, T.; Pooggin, M.M. Massive production of small RNAs from a non-coding region of *Cauliflower mosaic virus* in plant defense and viral counter-defense. *Nucleic Acids Res.* **2011**, *39*, 5003–5014.
52. Rajeswaran, R.; Aregger, M.; Zvereva, A.S.; Borah, B.K.; Gubaeva, E.G.; Pooggin, M.M. Sequencing of RDR6-dependent double-stranded RNAs reveals novel features of plant siRNA biogenesis. *Nucleic Acids Res.* **2012**, *40*, 6241–6254.
53. Aregger, M.; Borah, B.K.; Seguin, J.; Rajeswaran, R.; Gubaeva, E.G.; Zvereva, A.S.; Windels, D.; Vazquez, D.; Blevins, T.; Farinelli, L.; *et al.* Primary and secondary siRNAs in Geminivirus-induced gene silencing. *PLoS Pathog.* **2012**, *8*, e1002941.
54. Hany, U.; Adams, I.P.; Glover, R.; Bhat, A.I.; Boonham, N. The complete nucleotide sequence of Piper yellow mottle virus (PYMoV). *Arch. Virol.* **2013**, *158*, doi:10.1007/s00705-013-1824-2.
55. Quito-Avila, D.F.; Jelkmann, W.; Tzanetakis, I.; Keller, K.; Martin, R.R. Complete sequence and genetic characterization of *Raspberry latent virus*, a novel member of the family *Reoviridae*. *Virus Res.* **2011**, *155*, 397–405.
56. Ruiz-Ruiz, S.; Navarro, B.; Gisela, A.; Pena, L.; Navarro, L.; Moreno, P.; di Serio, F.; Flores, R. Citrus tristeza virus infection induces the accumulation of viral small RNAs (21–24 nt) mapping preferentially at the terminal region 3'- of the genomic RNA and affects the host small RNA profile. *Plant Mol. Biol.* **2011**, *75*, 607–619.
57. Loconsole, G.; Onelge, N.; Potere, O.; Giampetruzzi, A.; Bozan, O.; Satar, S.; de Stradis, A.; Savino, V.; Yokomi, R.K.; Saponari, M. Identification and characterization of *Citrus Yellow vein clearing virus*, a putative new member of the genus *Mandarivirus*. *Phytopathology* **2012**, *102*, 1168–1175.
58. Loconsole, G.; Saldarelli, P.; Doddapaneni, H.; Savino, V.; Martelli, G.P.; Saponari, M. Identification of a single-stranded DNA virus associated with citrus chlorotic dwarf disease, a new member of the family *Geminiviridae*. *Virology* **2012**, *432*, 162–172.
59. Maree, H.J.; Nel, Y.; Visser, M.; Coetzec, B.; Manicom, B.; Burger, J.T.; Rees, D.J.G. The study of plant virus disease etiology using next-generation sequencing technologies. In Proceedings of the 22nd International Conference on Virus and Other Transmissible Diseases of Fruit Crops, Rome, Italy, 3–8 June 2012; Abstract No. 48.
60. Yoshikawa, N.; Yamagishi, N.; Yaegashi, H.; Ito, T. Deep sequence analysis of viral small RNAs from a green crinkle-diseased apple tree. *Petria* **2012**, *22*, 292–297.
61. Candresse, T.; Marais, A.; Faure, C.; Carriere, S.; Gentit, P. Use of 454 pyrosequencing for the fast and efficient characterization of known or novel viral agents in *Prunus* materials. In Proceedings of the 22nd International Conference on Virus and Other Transmissible Diseases of Fruit Crops, Rome, Italy, 3–8 June 2012; Abstract No. 47.
62. Chiumenti, M.; Roberto, R.; Bottalico, G.; Campanale, A.; de Stradis, A.; Minafra, A.; Boscia, D.; Savino, V.; Martelli, G.P. Virus sanitation and deep sequence analysis of fig. In Proceedings of the 22nd International Conference on Virus and Other Transmissible Diseases of Fruit Crops, Rome, Italy, 3–8 June 2012; Abstract No. 163.

63. Thekke-Veetil, T.; Sabanadzovic, S.; Keller, K.E.; Martin, R.R.; Tzanetakis, I.E. Genome organization and sequence diversity of a novel blackberry Ampelovirus. In Proceedings of the 22nd International Conference on Virus and Other Transmissible Diseases of Fruit Crops, Rome, Italy, 3–8 June 2012; Abstract No. 191.
64. Candresse, T.; Marais, A.; Faure, C.; Gentit, P. Association of Little cherry virus 1 (LChV1) with the Shirofugen stunt disease and characterization of the genome of a divergent LChV1 isolate. *Phytopathology* **2013**, *103*, 293–308.
65. Roy, A.; Choudhary, N.; Guillermo, L.M.; Shao, J.; Govindarajulu, A.; Achor, D.; Wei, G.; Picton, D.D.; Levy, L.; Nakhla, M.K.; *et al.* A novel virus of the Genus *Cilevirus* causing symptoms similar to citrus leprosis. *Phytopathology* **2013**, *103*, 488–500.
66. Vives, M.C.; Velazquez, K.; Pina, J.A.; Moreno, P.; Guerri, J.; Navarro, L. Identification of a new enamovirus associated with citrus vein enation disease by deep sequencing of small RNAs. *Phytopathology* **2013**, *103*, 1077–1086.
67. Al Rwahnih, M.; Daubert, S.; Golino, D.; Rowhani, A. Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* **2009**, *387*, 395–401.
68. Coetzee, B.; Freeborough, M.J.; Maree, H.J.; Celton, J.M.; Rees, D.J.; Burger, J.T. Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. *Virology* **2010**, *400*, 157–163.
69. Pantaleo, V.; Saldarelli, P.; Miozzi, L.; Giampetruzzi, A.; Gisel, A.; Moxon, S.; Dalmay, T.; Bisztray, G.; Burgyan, J. Deep sequencing analysis of viral short RNAs from an infected Pinot noir grapevine. *Virology* **2010**, *408*, 49–56.
70. Zhang, Y.; Singh, K.; Kaur, R.; Qiu, W. Association of a novel DNA virus with the grapevine vein-clearing and decline syndrome. *Phytopathology* **2011**, *101*, 1081–1090.
71. Al Rwahnih, M.; Daubert, S.; U'rbez-Torres, J.R.; Cordero, F.; Rowhani, A. Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. *Arch. Virol.* **2011**, *156*, 397–403.
72. Giampetruzzi, A.; Roumi, V.; Roberto, R.; Malossini, U.; Yoshikawa, N.; la Notte, P.; Terlizzi, F.; Credi, R.; Saldarelli, P. A new grapevine virus discovered by deep sequencing of virus- and viroid-derived small small RNAs in Cv *Pinot gris*. *Virus Res.* **2012**, *163*, 262–268.
73. Al Rwahnih, M.; Dave, A.; Anderson, M.; Uyemoto, J.K.; Sudarshana, M.R. Association of a circular DNA virus in grapevine affected by red blotch disease in California. In Proceedings of the 17th Congress of ICVG, Davis, CA, USA, 7–14 October 2012.
74. Al Rwahnih, M.; Sudarshana, M.R.; Uyemoto, J.K.; Rowhani, A. Complete genome sequence of a novel Vitivirus isolated from grapevine. In Proceedings of the 17th Congress of ICVG, Davis, CA, USA, 7–14 October 2012.
75. Alabi, O.J.; Zheng, Y.; Jagadeeswaran, G.; Sunkar, R.; Naidu, R. High-throughput sequence analysis of small RNAs in grapevine (*Vitis vinifera* L.) affected by grapevine leafroll disease. In Proceedings of the 17th Congress of ICVG, Davis, CA, USA, 7–14 October 2012.
76. Poojari, S.; Alabi, O.J.; Fofanov, V.Y.; Naidu, R.A. A leafhopper-transmissible DNA virus with novel evolutionary lineage in the family *Geminiviridae* implicated in grapevine redleaf disease by next-Generation sequencing. *PLoS One* **2013**, *8*, e64194.

77. Navarro, B.; Pantaleo, V.; Gisel, A.; Moxon, S.; Dalmay, T.; Bistray, G.; di Serio, F.; Burgyan, J. Deep sequencing of viroid-derived small RNAs from grapevine provides new insight on the role of RNA silencing in plant-viroid interaction. *PLoS One* **2009**, *4*, e7686.
78. Di Serio, F.; Gisel, A.; Navarro, B.; Delgado, S.; Martínez de Alba, A.-E.; Donvito, G.; Flores, R. Deep sequencing of the small RNAs derived from two symptomatic variants of a chloroplastic viroid: Implications, for their genesis and for pathogenesis. *PLoS One* **2009**, *4*, e7539.
79. Di Serio, F.; Martínez de Alba, A.-E.; Navarro, B.A.; Gisel, A.; Flores, R. RNA-dependent RNA polymerase 6 delays accumulation and precludes meristem invasion of a viroid that replicates in the nucleus. *J. Virol.* **2010**, *84*, 2477–2489.
80. Martinez, G.; Donaire, L.; Llave, C.; Pallas, V.; Gomez, G. High-throughput sequencing of Hop stunt viroid-derived small RNAs from cucumber leaves and phloem. *Mol. Plant Pathol.* **2010**, *11*, 347–359.
81. Chiumenti, M.; Giampetruzzi, A.; Pirolo, C.; Morelli, M.; Saldarelli, P.; Minafra, A.; Bottalico, G.; la Notte, P.; Campanale, A.; Savino, V.; *et al.* Approaches of next generation sequencing to investigate grapevine diseases of unknown etiology. In Proceedings of the 17th Congress of ICVG, Davis, CA, USA, 7–14 October 2012.
82. Wu, Q.; Wang, Y.; Cao, M.; Pantaleo, V.; Burgyan, J.; Li, W.X.; Ding, S.W. Homology-independent discovery of replicating pathogenic circular RNAs by Deep sequencing and a new computational algorithm. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3938–3943.
83. Duan, Y.; Zhou, L.; Hall, D.G.; Li, W.; Doddapaneni, H.; Lin, H.; Liu, L.; Vahling, C.M.; Gabriel, D.W.; Williams, K.P.; *et al.* Complete genome sequence of citrus huanglongbing bacterium, “Candidatus Liberibacter asiaticus” obtained through metagenomics. *Mol. Plant Microbe Interact.* **2009**, *22*, 1011–1020.
84. Ng, T.F.F.; Duffy, S.; Polston, J.E.; Bixby, E.; Vallad, G.E.; Breitbart, M. Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. *PLoS One* **2011**, *6*, e19050.
85. Contaldo, N.; Canel, A.; Paltrinieri, S.; Bertaccini, A.; Nicolaisen, M. Phytoplasma detection and identification in grapevine by deep amplicon sequencing. In Proceedings of the 17th Congress of ICVG, Davis, CA, USA, 7–14 October 2012.
86. Nicolaisen, M.; Contaldo, N.; Makarova, O.; Paltrinieri, S.; Bertaccini, A. Deep amplicon sequencing reveals mixed phytoplasma infection within single grapevine plants. *Bull. Insectol.* **2011**, *64*, S35–S36.
87. Snyman, M.C.; Solofharivelo, M.C.; van der Walt, A.; Stephan, D.; Murray, S.; Burger, J.T. Deep sequencing analysis reveals modulated gene expression in response to aster yellow phytoplasma infection in *Vitis vinifera* cv. Chardonnay. In Proceedings of the 17th Congress of ICVG, Davis, CA, USA, 7–14 October 2012.
88. Palmano, S.; Saccardo, F.; Martini, M.; Ermacora, P.; Scortichini, M.; Abbà, S.; Marzachi, C.; Loi, N.; Firrao, G. Insights into phytoplasma biology through next generation sequencing. *J. Plant Pathol.* **2012**, *94*, 50–51.
89. Ehya, F.; Monavarfeshani, A.; Fard, E.M.; Farsad, L.K.; Nekouei, M.K.; Mardi, M.; Salekdeh, G.H. Phytoplasma-responsive microRNAs modulate hormonal, nutritional, and stress signalling pathways in Mexican lime trees. *PLoS One* **2013**, *8*, e6372.

90. Zhao, H.; Sun, R.; Albrecht, U.; Padmanabhan, C.; Wang, A.; Coffey, M.D.; Girke, T.; Wang, Z.; Close, T.J.; Roose, M.; *et al.* Small RNA profiling reveals phosphorous deficiency as a contributin factor in symptom expression for *Citrus Huanglongbing* disease. *Mol. Plant* **2013**, *6*, 301–310.
91. Paszkiewicz, K.; Studholme, D.J. *De novo* assembly of short sequence reads. *Brief Bioinform.* **2010**, *11*, 457–472.
92. Shendure, J.; Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145.
93. Zhang, W.; Chen, J.; Yang, Y.; Tang, Y.; Shang, J.; Chen, B. A practical application of *de novo* genome assembly software tools for next-generation sequencing technologies. *PLoS One* **2011**, *6*, e17915.
94. Horner, D.S.; Pavesi, G.; Castrignano, T.; D’Onorio de Meo, P.; Liuni, S.; Sammeth, M; Picardi, E.; Pesole, G. Bioinformatics approaches for genomics and post genomics applications of next generation sequencing. *Brief Bioinform.* **2009**, *11*, 181–197.
95. Metzker, M.L. Sequencing technologies—The next generation. *Nat. Rev. Genet.* **2010**, *11*, 31–46.
96. Miller, J.R.; Koren, S.; Sutton, G. Assembly algorithms for next generation sequencing data. *Genetics* **2010**, *95*, 315–327.
97. Pabinger, S.; Dander, A.; Fischer, M.; Snajder, R.; Sperk, M.; Efremova, M.; Krabichler, B.; Speicher, M.R.; Zschocke, J.; Trajanoski, Z. A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform.* **2013**, doi:10.1093/bib/bbs086.
98. *Bioinformatics for Biologists*; Pevzner, P., Shamir, R., Eds.; Cambridge University Press: Cambridge, UK, 2011; p. 392.
99. Wetterstrand, K.A. DNA sequencing costs: Data from the NHGRI Genome Sequencing Program (GSP). Available online: <http://www.genome.gov/sequencingcosts/> (accessed on 1 October 2013).
100. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63.
101. Hoffmann, C.; Minkah, N.; Leipzig, J.; Wang, G.; Arens, M.Q.; Tebas, P.; Bushman, F.D. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Res.* **2007**, *35*, e91.
102. Wang, C.; Mitsuya, Y.; Gharizadeh, B.; Ronaghi, M.; Shafer, R.W. Characterization of mutation spectra with ultra-deep pyrosequencing: Application to HIV-1 drug resistance. *Genome Res.* **2007**, *17*, 1195–1201.
103. Briese, T.; Paweska, J.T.; McMullan, L.K.; Hutchison, S.K.; Street, C.; Palacios, G.; Khristova, M.L.; Weyer, J.; Swanepoel, R.; Egholm, M.; *et al.* Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog.* **2009**, *5*, e1000455.
104. Barzon, L.; Lavezzo, E.; Militello, V.; Toppo, S.; Palù, G. Applications of next-generation sequencing technologies to diagnostic virology. *Int. J. Mol. Sci.* **2011**, *12*, 7861–7884.
105. Barba, M.; Hadidi, A. RNA silencing and viroids. *J. Plant. Pathol.* **2009**, *91*, 243–247.
106. Mlotshwa, S.; Pruss, G.J.; Vance, V. Small RNAs in viral infection and host Defense. *Trends Plant Sci.* **2008**, *13*, 375–382.
107. Liu, S.; Vijayendran, D.; Bonning, B.C. Next generation sequencing technologies for insect virus discovery. *Viruses* **2011**, *3*, 1849–1869.

108. Howell, W.E.; Thompson, D.; Scott, S. Virus-like disorders of fruit trees with undetermined etiology. In *Virus and Virus-Like Diseases of Pome and Stone Fruits*; Hadidi, A., Barba, M., Candresse, T., Jelkmann, W., Eds.; The American Phytopathological Society Press: St. Paul, MN, USA, 2011; pp. 259–265.
109. Martin, R.R. US Department of Agriculture, Agricultural Research Service, Corvallis, OR, USA. Personal communication, 2012.
110. *Virus Taxonomy Ninth Report of the International Committee on Taxonomy of Viruses*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier Academic Press: San Diego, CA, USA, 2012; p. 1327.
111. Hadidi, A.; Barba, M. Next-generation sequencing: Historical perspective and current applications in plant virology. *Petria* **2012**, *22*, 262–277.
112. Bar-Joseph, M.; Gera, A. On a revolutionary method for diagnosis of latent infections of viruses and the importance of its adoption to prevent the spread of citrus and other fruit tree diseases (in Hebrew). *Alon Hanotea* **2012**, *66*, 46–48.
113. Martelli, G.P. Grapevine virology highlights. In Proceedings of the 17th Congress of ICVG, Davis, CA, USA, 7–14 October 2012.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).